DOI: 10.1534/genetics.104.029231

Inferring the Mode of Speciation From Genomic Data: A Study of the Great Apes

Naoki Osada and Chung-I Wu[1]

Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637

Manuscript received March 23, 2004
Accepted for publication August 23, 2004

ABSTRACT

The strictly allopatric model of speciation makes definable predictions on the pattern of divergence, one of which is the uniformity in the divergence time across genomic regions. Using 345 coding and 143 intergenic sequences from the African great apes, we were able to reject the null hypothesis that the divergence time in the coding sequences (CDSs) and intergenic sequences (IGSs) is the same between human and chimpanzee. The conclusion is further supported by the analysis of whole-genome sequences between these species. The difference suggests a prolonged period of genetic exchange during the formation of these two species. Because the analysis should be generally applicable, collecting DNA sequence data from many genomic regions between closely related species should help to settle the debate over the prevalence of the allopatric mode of speciation.

THE allopatric mode of speciation is the tenet of the neo-Darwinian view of speciation (Mayr 1963). In this view, a geographical barrier preventing gene flow is a prerequisite for speciation. Without such barriers, gene exchanges during the process of species formation would obstruct the process as such exchanges would destroy the adaptive gene complexes and obliterate the accumulated differences between nascent species. On the other hand, there is no compelling population genetic reason why divergent adaptation cannot proceed in the presence of continuous gene flow (*e.g.*, Navarro and Barton 2003). A most common mode may be parapatric speciation when nascent species are geographically connected by gene flow (Mayr 1963; Endler 1977). The extreme form of gene flow is represented by sympatric speciation (Dieckmann and Doebeli 1999; Kondrashov and Kondrashov 1999). Parapatric speciation may best be envisioned at the genic level (Wu and Ting 2004) where portions of the genome progressively become divergently adapted and hence nonexchangeable between nascent species. The genealogical history of the genome would therefore be mosaic with disparate divergence time among different loci. While most previous tests of the allopatric *vis-à-vis* parapatric mode of speciation were based on ecological or biogeographical considerations (Endler 1977; Butlin 1998; Coyne and Price 2000), only a few studies have utilized multiple DNA sequences (Kliman *et al.* 2000; Machado *et al.* 2002) for this purpose. This analysis represents a genome-wide perspective on the same issue.

[1]*Corresponding author:* Department of Ecology and Evolution, University of Chicago, 1101 E. 57th St., Chicago, IL 60637.
E-mail: ciwu@uchicago.edu

In strict allopatry, all the genes in the genome should have the same divergence history (t in Figure 1A) but vary in the coalescence time, which is exponentially distributed with mean equal to $2N_e$ (N_e being the effective population size at the time of speciation, Figure 1A). We discuss more complex forms of allopatric speciation later. Note that time is measured in units of generation throughout this report. A large variance in DNA divergence can be due to either variation in t across loci or a larger-than-estimated N_e, both of which can enhance the variance in divergence among loci. Although there are many studies for estimating N_e, they all assume constant t across loci, or strict allopatry, precisely what we wish to test. Interestingly, when t is assumed to be a constant, the estimated N_e's for the ancient species are usually far larger than those for the extant populations (Ruvolo 1997; Takahata and Satta 1997; Chen and Li 2001; Yang 2002; Wall 2003). These studies thus hint at the possibility of nonconstant t.

In this study, we compare coding and intergenic regions for their evolutionary dynamics during speciation. In allopatry, these two types should have the same dynamics but, under the parapatric model of speciation, could have very different histories. Figure 1B illustrates this point, on the assumption that coding sequences are more likely than intergenic sequences to be associated with hybrid incompatibility or differential adaptation. The potential for coding regions to successfully move across nascent species boundaries may be curtailed early (Figure 1B). On the other hand, intergenic regions, experiencing less impediment to their trafficking between nascent species, should continue to be exchangeable until openings in the reproductive barrier are completely sealed. This contrast has been reported for Dro-

sophila between DNA sequences at or near a speciation gene (Ting *et al.* 2000). A recent report also assumes that the common ancestors of human and chimpanzee went through a period of parapatry (Navarro and Barton 2003). However, the observations were reanalyzed in light of outgroup data and were suggested to result from events unrelated to speciation (Lu *et al.* 2003; Navarro *et al.* 2003).

MATERIALS AND METHODS

Sequence data: We collected 98 common chimpanzee (*Pan troglodytes*) sequences from the GenBank database, 93 from the 5′-conseusus sequences of Sakate *et al.* (2003), 19 newly determined full-length cDNA sequences from Ryuichi Sakate and Momoki Hirai (University of Tokyo), and 135 genomic sequences of chimpanzee chromosome 22 corresponding to Ensembl genes of human chromosome 21. Seventy-six gene sequences of gorilla were collected from GenBank. We removed from our analysis MHC sequences, whose genealogy was deeper than human-chimpanzee divergence due to strong balancing selection (Satta *et al.* 1999). DDBJ/EMBL/GenBank accession numbers of newly determined sequences are AB188273–AB188288.

The sequences were aligned by using the ClustalW program (Thompson *et al.* 1994) and corrected by visual inspection. Numbers of synonymous and nonsynonymous substitutions were estimated by the method of Li (1993) with equal weighting among pathways for multiple substitutions in a codon (Pamilo and Bianchi 1993). Fifty-three intergenic sequences were obtained from Chen and Li (2001). Ninety pairs of 2-kb intergenic sequences of human and chimpanzee, which are at least 10 kb apart from genic regions annotated by Ensembl, were obtained from genomic sequences of human chromosome 21 and chimpanzee chromosome 22. Numbers of substitutions of intergenic regions were estimated by using Kimura's two-parameter method (Kimura 1980).

Maximum-likelihood estimate of divergence time and ancestral population size: We designate k_i as the number of synonymous changes for the ith sequence [either coding sequence (CDS) or intergenic sequence (IGS)]. The probability of observing k_i is given by

$$P(k_i) = \frac{e^{-l_i\tau_i}}{1 + l_i\theta_i} \sum_{d=0}^{k_i} \frac{(l_i\tau_i)^d}{d!}\left(\frac{l_i\theta_i}{1 + l_i\theta_i}\right)^{k_i-d}, \quad (1)$$

where $\tau_i = 2tu_i$ and $\theta_i = 4N_eu_i$ (Equation 5 in Takahata and Satta 1997). l_i is the length of sequence i and u_i is the per-nucleotide substitution rate for the ith sequence. Equation 1 has two components—the Poisson distribution in the divergence portion ($\tau_i = 2tu_i$) and the "mismatch distribution" in the coalescence portion ($\theta_i = 4N_eu_i$), where the absence of intragenic recombination is assumed.

Without the outgroup: We first assume that the substitution rate for CDS (and separately for IGS) is uniform across m loci when the outgroup sequences are not available for calibrating the variation in the mutation rate. Let $\tau = 2tu$ and $\theta = 4N_eu$. The log-likelihood for Equation 1 becomes

$$L(\tau, \theta) = \ln \prod_{i=1}^{m}\left[\frac{e^{-l_i\tau}}{1 + l_i\theta} \sum_{d=0}^{k_i} \frac{(l_i\tau)^d}{d!}\left(\frac{l_i\theta}{1 + l_i\theta}\right)^{k_i-d}\right].$$

The maximum-likelihood estimates (MLEs) of τ and θ were found by numerical iteration.

With the outgroup: We now use sequences from an outgroup species, say the orangutan, to filter out the variation in u_i. Let the divergence time between human and the outgroup be T and the substitution number of the ith locus between these two species be K_i. We assume that $K_i = 2l_iTu_i$ without considering the coalesence component because, if T is sufficiently large, the impact of $2N_e$ should be insignificant. Noting that the divergence time between human and chimpanzee is t, we can now replace τ_i and θ_i with $(t/T)(K_i/l_i) = \alpha(K_i/l_i)$ and $(2N_e/T)(K_i/l_i) = \beta(K_i/l_i)$, respectively, in Equation 1. $\alpha = t/T$ and $\beta = 2N_e/T$ are the two parameters to be estimated by MLE.

The log-likelihood function is

$$L(\alpha, \beta|2l_iTu_i = K_i) = \ln \prod_{i=1}^{m}\left[\frac{e^{-\alpha K_i}}{1 + \beta K_i} \sum_{d=0}^{k_i} \frac{(\alpha K_i)^d}{d!}\left(\frac{\beta K_i}{1 + \beta K_i}\right)^{k_i-d}\right]$$

and the MLE of the two parameters, α and β, can be found by numerical iteration.

Computer simulations: In the allopatric model, the divergence of all genes is fixed at t (Figure 1A), while in the parapatric model (Figure 1B), the divergence time is uniformly distributed between t and $2t$. For 1000 loci, random integers in the range of 500–1000 are generated for the number of sites. Coalescent times are generated as exponential distribution whose mean is $2N_e$ while N_e satisfies $\gamma = t/2N_e = 10$. t corresponds to 1% nucleotide divergence with 1.5×10^{-8} substitutions per generation per site. The number of substitutions is assigned according to the Poisson distribution where the mutation rate is uniform among loci. A confidence interval of 95% was calculated by 1000 iterations.

Congruence between gene genealogy and species phylogeny: Seventy-six gene sequences of human, chimpanzee, gorilla, and orangutan were used for the congruence test. The observed genealogies can be (M, M, m), (m, M, M), or (M, m, M) (see Figure 1C), where m and M are the ancestral and derived variants, respectively. (m, M, M) and (M, m, M), where gorilla shares the derived variant with either chimpanzee or human, are incongruent with the species phylogeny, in which human and chimpanzee are the closest relatives. In our analysis, we first treated each site independently. All CG to TG and CG to CA substitutions were masked from the analysis because of the very high rate of changes at such CpG sites, which often results in genealogies incongruent with the species phylogeny. Adjacent variant sites within the same locus that show the same genealogical pattern are counted as one segment. A locus may have more than one segment showing different phylogenetic patterns, presumably due to recombination.

We designate the observed number of segments that show the pattern of (M, M, m), (m, M, M), and (M, m, M) a_1, b_1, and c_1, respectively, for IGS. Likewise, the numbers are a_2, b_2, and c_2, respectively, for CDS. Under the null hypothesis that $t'/2N_e'$ is the same between coding and intergenic regions (Figure 1C), the likelihood ratio R is

$$R = \left(\frac{a_1}{n_1}\right)^{a_1}\left(\frac{b_1 + c_1}{2n_1}\right)^{b_1+c_1}\left(\frac{a_2}{n_2}\right)^{a_2}\left(\frac{b_2 + c_2}{2n_2}\right)^{b_2+c_2} \Bigg/ \left(\frac{a_3}{n_3}\right)^{a_3}\left(\frac{b_3 + c_3}{2n_3}\right)^{b_3+c_3},$$

where $a_3 = a_1 + a_2$, $b_3 = b_1 + b_2$, $c_3 = c_1 + c_2$, $n_1 = a_1 + b_1 + c_1$, $n_2 = a_2 + b_2 + c_2$, and $n_3 = a_3 + b_3 + c_3$ (derived from Equation 7 in Wu 1991).

RESULTS AND DISCUSSION

We estimate $\tau = 2tu$ and $\theta = 4N_eu$, where u is the per-nucleotide substitution rate, by the maximal-likelihood (ML) method (Takahata and Satta 1997; see materials and methods). Our objective is to test if t is the same between coding and intergenic regions. However, because u many not be the same between two regions,

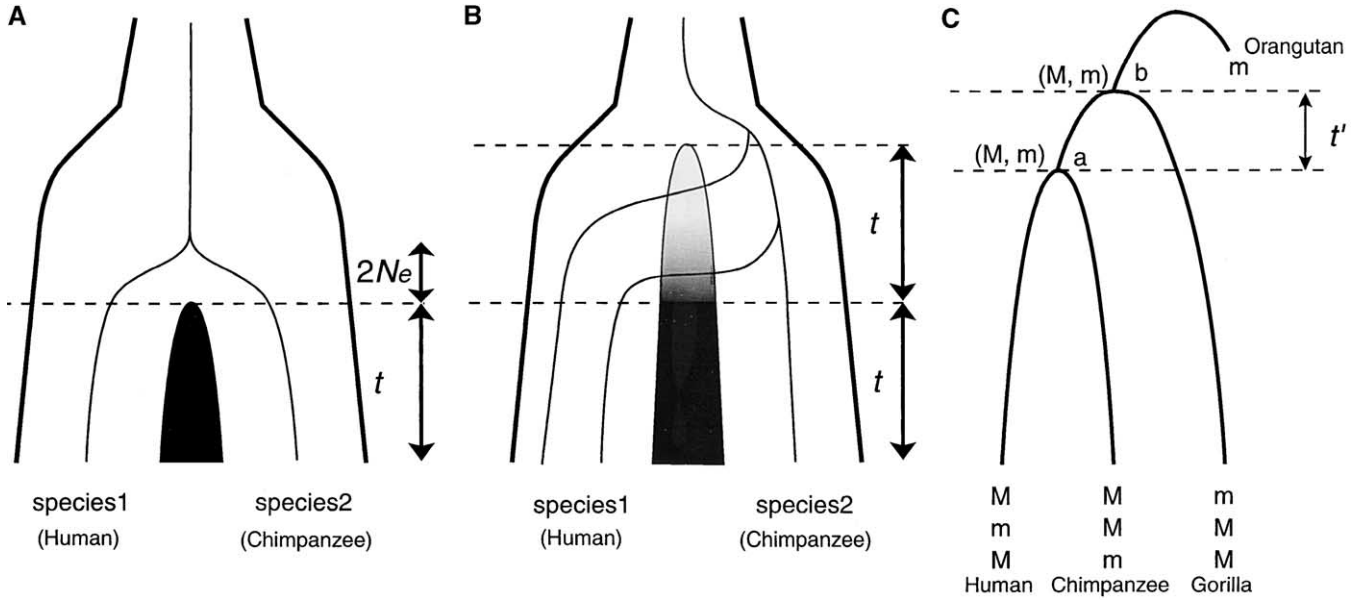

FIGURE 1.—(A) Allopatric speciation. In strict allopatry, there is no gene flow beyond the time of separation. All genes hence have diverged for a fixed time t and further coalesce with an average length of $2N_e$ generations. (B) Parapatric speciation. Under the parapatric model, there is a period of time when gene flow between nascent species is possible. The intensity of shade indicates the strength of the barrier to gene flow. For genomic regions (such as CDSs) associated with reproductive incompatibility, early cessation of gene flow is likely. For regions free of such association (including most IGSs), gene flow may continue until relatively late. (C) Segregation of polymorphisms (m for the ancestral and M for the derived variant) under the allopatric model. The two speciation events, denoted a and b, were separated by t', during which time the effective population size is N_e'.

we define $\gamma = \tau/\theta = t/2N_e$ and test if γ is the same between the two regions. γ is the relative divergence accrued after, *vis-à-vis* before, speciation and should be constant under the null hypothesis of allopatry.

To know how parapatry might affect the estimation of γ when allopatry (*i.e.*, constant t across loci) is incorrectly assumed, we carried out computer simulations. In the simplest case, the divergence time in the allopatric model is fixed at t (Figure 1A), while the divergence time in the parapatric model is uniformly distributed between t and $2t$ (Figure 1B). More complex simulations have been done but the results can be qualitatively stated as such: parapatry generally results in the underestimation of γ ($= t/2N_e$). Even when the true γ is 50% larger in parapatry than in allopatry, as in the case of Figure 1, the estimated numbers are nevertheless in the opposite direction (Table 1). The reason for this seemingly paradoxical result is that, under parapatry, the estimate of $2N_e$ is greatly inflated to account for the variation in the level of divergence among loci. Hence, we expect γ to be underestimated when allopatry is incorrectly imposed on data that have a variable divergence time. Coding sequences probably fit this characterization better than intergenic sequences (Figure 1B).

TABLE 1

Simulation results from the schemes of Figure 1, A (allopatric) and B (parapatric)

	Divergence time[a]	γ (expected)	95% C.I. of γ (estimated)
Model A (allopatric)	t	10	$7.36 \sim 18.54$
Model B (parapatric)	$t \sim 2t$	15	$3.48 \sim 4.72$

γ was estimated from 100 rounds of simulations. The parameter values of t, N_e, and u are given in MATERIALS AND METHODS.

[a] Except for coalescence time.

We used 345 CDSs and 143 IGSs from human and chimpanzee and conducted the likelihood-ratio test between the two hypotheses, $\gamma_{CDS} = \gamma_{IGS} = \gamma_0$ and $\gamma_{CDS} \neq \gamma_{IGS}$, where γ_{CDS} and γ_{IGS} are MLEs for the CDS and IGS, respectively (TAKAHATA and SATTA 1997). Under the null hypothesis, the MLE of γ_0 is 1.89 and the log-likelihood value is -1098.588 (Table 2). Under the alternative hypothesis, the MLEs for the two regions are $\gamma_{CDS} = 1.31$ and $\gamma_{IGS} = 2.45$ and the log-likelihood value is -1096.226 (Table 2). The likelihood-ratio test between the two models yields a significant result ($P = 0.027$). Because the variation among loci in the number of CpG sites, which exhibit high mutability (HELLMANN *et al.* 2003), may have an impact on our estimation, we reestimated γ by masking all CG to TG and CG to CA substitutions. The likelihood-ratio test leads to the same conclusion ($P = 0.006$, see supplementary Table 1 at http://www.genetics.org/supplemental/). Strictly speaking, because N_e may be smaller for the coding than for the intergenic region, as the former is generally less variable than the latter (PLUZHNIKOV *et al.* 2002), the null hypothesis should be $\gamma_{IGS} \leq \gamma_{CDS}$, making our test conservative. The null hypothesis of $\gamma_{IGS} \leq \gamma_{CDS}$ is thus rejected.

For the method to be of general use in testing allopatric speciation, the need for DNA sequences should not

TABLE 2

Estimation of $\tau = 2tu$ and $\theta = 4N_e u$ (see Figure 1A) in pairwise comparisons among human, chimpanzee, and gorilla ($\gamma = t/2N_e$)

	Human-chimpanzee ($n_C = 345$, $n_I = 143$)	Human-gorilla ($n_C = 76$, $n_I = 53$)	Chimpanzee-gorilla ($n_C = 76$, $n_I = 53$)
H_0: $\gamma_{CDS} = \gamma_{IGS} = \gamma_0$			
τ_{CDS}	0.00855	0.01299	0.01317
θ_{CDS}	0.00454	0.00500	0.00380
τ_{IGS}	0.00876	0.01094	0.01204
θ_{IGS}	0.00466	0.00421	0.00347
γ_0	1.88	2.60	3.47
ln L	−1093.971	−276.117	−270.641
H_1: $\gamma_{CDS} \neq \gamma_{IGS}$			
τ_{CDS}	0.00748	0.01286	0.01112
θ_{CDS}	0.00579	0.00514	0.00618
γ_{CDS}	1.29	2.50	1.80
θ_{IGS}	0.00936	0.01099	0.01300
θ_{IGS}	0.00382	0.00414	0.00242
γ_{IGS}	2.45	2.65	5.37
ln L	−1091.530	−276.114	−269.906
	$P = 0.027$	$P = 0.950$	$P = 0.224$

exceed what we used above. A need for >500 sequences would make the method impractical for most specie pairs. Nevertheless, between human and chimpanzee, 7645 orthologous sequences are available (Clark *et al.* 2003) to back up the above analysis. For this large dataset, γ_{CDS} is 1.20, which leads to an even more significant likelihood ratio ($P = 0.0003$, see supplementary Table 1). Above 500 sequences, an increase in sample size >500 in this case appears to yield a diminishing return.

To standardize the divergence measure and make it independent of the underlying mutation rate, we also calibrate the human-chimpanzee divergence against the divergence between these two species and an outgroup. We were able to use only 76 CDSs and 53 IGSs from human, chimpanzee, and orangutan for this purpose. It is assumed that the level of divergence between human and orangutan is a function of their divergence time, T, without much influence by the ancestral polymorphism, the contribution of which should be relatively small here. The key parameters are now $\alpha = t/T$ and $\beta = 2N_e/T$ (see Figure 1 and materials and methods). By doing so, γ ($= \alpha/\beta$) was estimated to be 1.55 and 37.3 for the CDSs and IGSs, respectively. While the estimates are different from those of Table 2 due to both the small sample sizes and the inherent variability in the estimation of γ (see Table 1), the general trend of $\gamma_{IGS} \gg \gamma_{CDS}$ is observed.

When calibrated against the divergence from the orangutan, the divergence in CDS and IGS between human and chimpanzee can in fact be directly compared since the governing parameters, $\alpha = t/T$ and $\beta = 2N_e/T$, depend only on the common elements, t, T, and $2N_e$. For each locus, we therefore compute the relative divergence $d_R = d_{hc}/[(d_{ho} + d_{co})/2]$, where d_{hc}, d_{ho}, and d_{co} are the levels of divergence between human and chimpanzee, human and orangutan, and chimpanzee and orangutan, respectively. The mean of d_R is 0.522 for CDS and 0.404 for IGS ($P = 0.030$) while the variance of d_R is 0.166 for CDS and 0.037 for IGS ($P < 10^{-7}$). The results suggest that, on average, coding regions have deeper genealogy than intergenic regions and the variation is larger in the former than in the latter, as hypothesized in Figure 1.

The analysis of Table 2 has also been applied to the divergence between gorilla and either human or chimpanzee (node b of Figure 1C). By using 76 coding and 53 intergenic sequences the null hypothesis of allopatry cannot be rejected ($P = 0.950$ for human-gorilla and $P = 0.224$ for chimpanzee-gorilla). Although the results are not significant, the chimpanzee-gorilla comparison appears to be very different from the human-gorilla divergence. In the former, $\gamma_{IGS} > \gamma_{CDS}$ and the difference is larger than that in the human-chimpanzee comparison (Table 2). Given the small number of sequences from gorilla, there is little statistical power to resolve the issue at this moment. Nevertheless, chimpanzee and gorilla occupy mainly western Africa, whereas ecological and paleontological evidence suggests proto-humans have migrated to eastern and southern Africa (Leakey *et al.* 2001). Hence a prolonged period of gene flow between ancestral chimpanzee and gorilla seems plausible.

Finally, we may analyze the joint effect of two speciation events in succession, as shown in Figure 1C. We assume that the species phylogeny of Figure 1C is strictly correct and the two allopatric events are separated by time t' during which the effective population size was

TABLE 3

Number of DNA segments that support any of the three phylogenetic patterns—(HC)(GO), (CG)(HO), or (HG)(CO), where humans (H), chimpanzees (C), and gorillas (G) and orangutans (O) share the variant with one other species only ($P = 0.013$)

	(HC)(GO)	(CG)(HO)	(HG)(CO)
Intergenic ($n = 53$)	23 (63.9%)	6 (16.7%)	7 (19.4%)
Coding ($n = 76$)	26 (49.1%)	14 (26.4%)	13 (24.5%)

N'_e. The probability of having a genealogy incongruent with the species phylogeny, (m, M, M) or (M, m, M) of Figure 1C, is a function of $t'/2N'_e$ (Nei 1987; Wu 1991). The null hypothesis, again, is that $t'/2N'_e$ is the same for coding and intergenic regions. We used 53 intergenic and 76 coding sequences from human (H), chimpanzee (C), gorilla (G), and orangutan (O). Orangutan is used as an outgroup to distinguish the derived mutation, M, from the ancestral state, m. We masked all substitutions at CpG sites and then classified the patterns of independently segregating sites into the three categories shown in Figure 1C.

The proportion of incongruent genealogies is 0.509 and 0.361 for CDS and IGS, respectively (Table 3). The result of the likelihood-ratio test is not significant ($P = 0.166$), probably due to the small number of sequence fragments. With a larger sample size, say, twice the number of genes in Table 3, this approach should be useful for addressing the issue of allopatric speciation.

By analyzing the divergence among hundreds of DNA sequences, we inferred that the speciation history between human and chimpanzee cannot be the same for coding and intergenic regions. Genomic sequences between closely related species may thus provide new opportunities to settle the debate on the prevalence of allopatric speciation. In a series of analyses, Hey, Wakeley, and colleagues (Wakeley and Hey 1997; Kliman *et al.* 2000; Machado *et al.* 2002) addressed the same problem of parapatry using both the polymorphism and divergence data. While their approach utilizes more information per locus, we believe the approach outlined here will be more practical for several reasons. First, in the immediate future, there will be a torrent of data consisting of one sequence per gene per species. Second, polymorphism data will not be useful for resolving the mode of speciation in many species—human *vs.* chimpanzee being an obvious example. Third, the effect of selection on polymorphism can be more difficult to gauge than that on divergence, making the inference on speciation more difficult.

Finally, allopatric speciation could have more complex patterns than portrayed here. It may happen between deeply subdivided but connected populations where disparate genealogies preexisted when speciation took place allopatrically. Such a model can be seen as a hybrid between parapatry and allopatry. However, if populations can evolve to become differentially adapted and strongly subdivided in the presence of gene flow, it seems plausible that they can continue to diverge without a newly erected geographical barrier to stop gene flow completely. Moreover, the restriction of gene flow imposed by the diverging genomes should continue to strengthen as incompatibilities evolve to encompass larger and larger linkage blocks (Wu and Ting 2004). Testing such a hybrid model may require both the divergence and polymorphism data at the genomic level (Wakeley and Hey 1997; Kliman *et al.* 2000; Machado *et al.* 2002). At this moment, testing strict (and simple) allopatry among diverse taxa, as outlined here, seems a logical first step.

We thank T. Nagylaki, H. Tang, H. Y. Wang, J. Lu, and Y.-X. Fu for providing theoretical advice and/or helping with data analysis; F. C. Chen and W. H. Li for kindly providing the intergenic sequence data; R. Sakate and M. Hirai for the chimpanzee coding sequences; K. Hashimoto and C. K. J. Shen for the macaque cDNA sequences; and J. Shapiro, M. Kohn, B. Harr, M. Long, L. Zhang, I. Boussy, and J. Spofford for comments and discussions.

LITERATURE CITED

Butlin, R., 1998 What do hybrid zones in general, and the *Chorthippus parallelus* Zone in particular, tell us about speciation?, pp. 367–378 in *Endless Forms: Species and Speciation*, edited by D. Howard and S. Berlochers. Oxford University Press, Oxford.

Chen, F.-C., and W.-H. Li, 2001 Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees Am. J. Hum. Genet. **68:** 444–456.

Clark, A. G., S. Glanowski, R. Nielsen, P. D. Thomas, A. Kejariwal *et al.*, 2003 Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science **302:** 1960–1963.

Coyne, J., and T. D. Price, 2000 Little evidence for sympatric speciation in island birds. Evolution **54:** 2166–2171.

Dieckmann, U., and M. Doebeli, 1999 On the origin of species by sympatric speciation. Nature **400:** 354–357.

Endler, J. A., 1977 Geographic variation, speciation, and clines. Monogr. Popul. Biol. **10:** 1–246.

Hellmann, I., S. Zollner, W. Enard, I. Ebersberger, B. Nickel *et al.*, 2003 Selection on human genes as revealed by comparisons to chimpanzee cDNA. Genome Res. **13:** 831–837.

Kimura, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman *et al.*, 2000 The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics **156:** 1913–1921.

Kondrashov, S., and F. A. Kondrashov, 1999 Interactions among quantitative traits in the course of sympatric speciation. Nature **400:** 351–354.

Leakey, M. G., F. Spoor, F. H. Brown, P. N. Gathogo, C. Kiarie *et al.*, 2001 New hominin genus from eastern Africa shows diverse middle Pliocene lineages. Nature **410:** 433–440.

Li, W.-H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36:** 96–99.

Lu, J., W.-H. Li and C.-I Wu, 2003 Comment on chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. Science **302:** 988.

Machado, C. A., R. M. Kliman, J. A. Markert and J. Hey, 2002 In-

ferring the history of speciation from multilocus DNA sequence data: the case of Drosophila pseudoobscura and close relatives. Mol. Biol. Evol. **19:** 472–488.

Mayr, E., 1963 *Animal Species and Evolution*. Belknap Press, Cambridge, MA.

Navarro, A., and N. H. Barton, 2003 Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. Science **300:** 321–324.

Navarro, A., T. Marques-Bonet and N. H. Barton, 2003 Response to comment on chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. Science **302:** 988.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Pamilo, P., and N. O. Bianchi, 1993 Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol. Biol. Evol. **10:** 271–281.

Pluzhnikov, A., A. D. Rienzo and R. R. Hudson, 2002 Inferences about human demography based on multilocus analyses of noncoding sequences. Genetics **161:** 1209–1218.

Ruvolo, M., 1997 Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. Mol. Biol. Evol. **14:** 248–265.

Sakate, R., N. Osada, M. Hida, S. Sugano, I. Hayasaka *et al.*, 2003 Analysis of 5′-end sequences of chimpanzee cDNAs. Genome Res. **13:** 1022–1026.

Satta, Y., H. Kupfermann, Y. J. Li and N. Takahata, 1999 Molecular clock and recombination in primate Mhc genes. Immunol. Rev. **167:** 367–379.

Ting, T., S. C. Tsaur and C.-I Wu, 2000 The phylogeny of closely related species as revealed by the genealogy of a speciation gene, Odysseus. Proc. Natl. Acad. Sci. USA **97:** 5313–5316.

Takahata, N., and Y. Satta, 1997 Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. Proc. Natl. Acad. Sci. USA **94:** 4811–4815.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Wakeley, J., and J. Hey, 1997 Estimating ancestral population parameters. Genetics **145:** 847–855.

Wall, J. D., 2003 Estimating ancestral population sizes and divergence times. Genetics **163:** 395–404.

Wu, C.-I, 1991 Inferences of species phylogeny in relation to segregation of ancient polymorphism. Genetics **127:** 429–435.

Wu, C.-I, and C. T. Ting, 2004 Genes and speciation. Nat. Rev. Genet. **5:** 114–122.

Yang, Z., 2002 Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. Genetics **162:** 1811–1823.

Communicating editor: Y.-X. Fu